Real Time Gender Detection of The Speech Signal Using Auto Correlation

A Sai Abhishek[1], Krishna Prakash[2]

Abstract- *There are a lot of differences between Male and Female voices. We as human beings have the ability to distinguish between a male and a female voice, and we can do it in an instant. This difference is biological. Male and female voices sound differently even at the same pitch. This is due to the variation in the size and shape of the vocal cords and the larynx between men and woman. As humans we have the natural skill to determine the gender of a person as soon as we hear the voice. In this paper, we will be looking at an efficient algorithm to identify the gender of the speaker by analysing the speech signal.*

Keywords- Gender Identification, speech analysis, signal processing, correlation, fundamental frequency.

I. INTRODUCTION

It is evident that there are a lot of differences between male and female voices. One obvious difference is the pitch. Men, on an average, tend to speak almost an octave lower than their counterparts. Hence, men have a low-pitched voice, while women on the other hand, have a higher pitched voice. Human ears are capable of picking up a wide variety of sounds having different frequencies, pitches, depth etc, and our brains process it and interpret these signals and send back electrical impulses.

The ability to detect pitch varies from one animal to another. The Basilar membranes in the inner ear contain hair cells that react differently to different frequencies. They are naturally frequency selective. The membrane is located in the cochlea of the inner ear, where the movement of the cochlear fluid causes it to vibrate. The organ of Corti is the structure located on the basilar membrane of the inner ear that contains the auditory receptors. The hair cells are spread out according to how profound the frequency is that exhilarates it.

II. THEORY

The main principle used in this paper to determine the gender is **Correlation**. Correlation is the process of determining the similarities between two signals. In a way, correlation describes the mutual relationship between these signals. Correlation indicates the similarity between two signals. This works when we consider the voice signal also. So basically, if we want to check how much similarity exists between the two signals, correlation can be used. Correlation generally is of two types, Autocorrelation and Cross-correlation.

2.1 Auto Correlation

Auto-correlation is used when you correlate a given signal with the delayed version of the same signal. In a way, auto-correlation is also cross-correlation but with a delayed or lagged version of the same signal. The mathematical expression for auto-correlation is as follows.

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)\, x^{\star}(t-\tau)\, dt \tag{1}$$

Fig1. Mathematical Expression for auto-correlation

The equation above is used for continuous time signal. The * denotes the complex conjugate.

2.2 Cross-correlation

This is a type of correlation where one signal is correlated with a completely different signal to know how much resemblance exists between them. The mathematical expression for cross-correlation is as follows.

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)\, y^{\star}(t-\tau)\, dt \tag{2}$$

Fig 2. Mathematical Expression for Cross-correlation

We can also find the correlation between discrete time signals, but since our focus is on human speech signal which is a continuous time signal, we will be sticking only to continuous time signals.

2.3 Cepstrum

Cepstrum is a method used in homomorphic signal processing, to convert the convoluted signal into the sums of their cepstra. Power Cepstrum is widely used as a feature to represent the human and musical signals. Power Cepstrum is a very well-known method to detect the pitch of a voice signal. Mathematically, cepstrum is the result of inverse Fourier Transform of the logarithm of the signal spectrum. Cepstrum analysis is a type of non-linear signal processing technique. Its mathematical expression is as follows.

$$\hat{x} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log[X(e^{j\omega})]e^{j\omega n}\,d\omega. \tag{3}$$

Fig 3. Mathematical Expression for Cepstrum Analysis

2.4 Fundamental Frequency

Human speech has a frequency range between 12Hz to 8Khz. Different areas of this particular frequency spectrum have different implications and functions. For example, the lowest parts of the frequency spectrum of a raw speech signal helps to identify the pitch of the voice. The lowest frequency of the human speech is referred to as the fundamental frequency. As it is the lowest frequency, it also serves the purpose of giving us information about pitch of the voice.

The fundamental frequency range for the human speech typically ranges between 125Hz to 200Hz. As we discussed above, we established the fact that women typically have a higher pitch, and men have a lower pitch. By looking at the frequency range we can say that 160hz is close the mid-point of the spectrum. Generally, women have fundamental frequency which is more than the mid-point or 160Hz, and men have a fundamental frequency less than 160Hz.

If F0 is the fundamental frequency then the length of a single period in seconds is:

$$T = \frac{1}{F_0}\,. \tag{4}$$

This means that the speech signal repeats itself after a time period of T seconds.

Hence, the best and the simplest way of representing the fundamental frequency is to repeat the signal after a delay of T seconds. Consider the sampling frequency to be Fs. Then if the signal repeats itself after a delay of L samples,

$$L = F_s T = \frac{F_s}{F_0}\,. \tag{5}$$

Where Fs is sampling frequency and F0 is the fundamental frequency.

III. WORKING AND IMPLEMENTATION

3.1 Flowchart

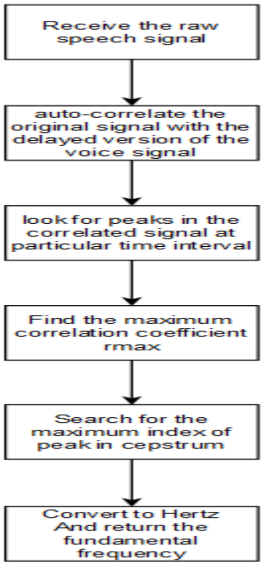

Receive the raw speech signal
auto-correlate the original signal with the delayed version of the voice signal
look for peaks in the correlated signal at particular time interval
Find the maximum correlation coefficient rmax
Search for the maximum index of peak in cepstrum
Convert to Hertz And return the fundamental frequency

Fig 4. Flowchart of the Algorithm

3.2 Implementation

The main problem with finding the fundamental frequency is to take a portion of the signal and to find the more frequently repeated frequency. Since not all signals are periodic, and even if they are periodic, their fundamental frequency can be varying if there are multiple signals, and if the raw signal is contaminated with noise, difficulties may arise.

Hence, the most reliable way to attain the most dominant frequency in a speech signal is to use the *cepstrum.* To obtain an accurate value of the fundamental frequency we look for peaks in the frequency region. That is why we use auto-correlation; to find where the peaks are located.

We have implemented this algorithm using MATLAB.

3.4 Working

Firstly, we record our speech signal for 10 seconds. This is the raw speech signal that we will be finding the fundamental frequency for. Once we get the raw signal, we need to define the part of signal that we will be correlating with. In our case, we have taking minimum frequency at 50Hz and maximum frequency of 1000Hz. We have set the Sampling Frequency Fs at 16000Hz.

Then we perform autocorrelation of the delayed version of the raw signal with the original version of the speech signal. After correlation process, if we plot the graph you can see that the autocorrelation functions has maximum peak at t=0, and we also notice that at other intervals also we have a few peaks. Therefore, now we can accurately estimate the frequency by looking for peaks at the corresponding frequency range.

Once, we find at which index we are having the maximum peaks, we need to convert that index into Hertz, to get the final fundamental frequency. For example, we can take the values between 1ms and 20ms and find the index for that particular range. Then by using the sampling frequency defined we convert the time into frequency. After this step we have successfully obtained what should be the fundamental frequency of this speech signal.

Our final objective is to determine the gender of the speaker, so we shouldn't stop with determining the fundamental frequency. We need to find a way to use the fundamental frequency to accurately identify the gender. As we have discussed in the theory part, we have established the fact that women have higher frequency typically more than 160Hz, and men have lower pitch compared to women and that generally falls under the 160Hz threshold. Hence, it makes sense to set the threshold value at 160Hz.

Using a simple conditional statement, we can conclude that if the fundamental frequency is obtained fundamental frequency is less than 160Hz it's a male voice, and if its more than the threshold level, it should be a female voice.

IV. RESULTS

4.1 Case 1: Male Voice:

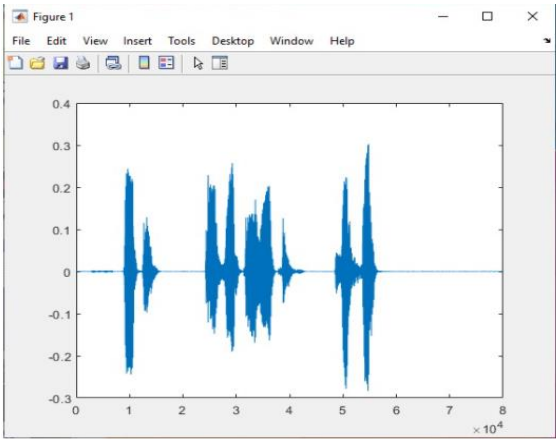


Fig 5. Raw Speech signal of a male voice

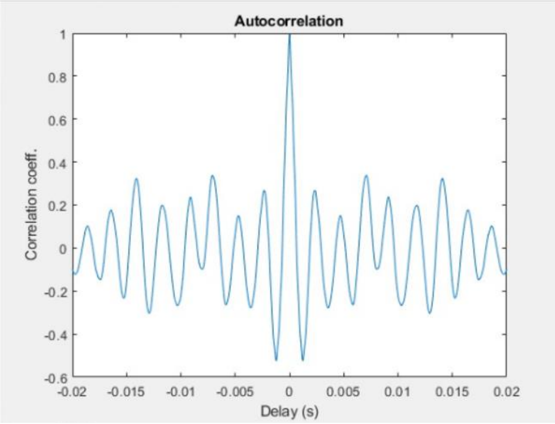
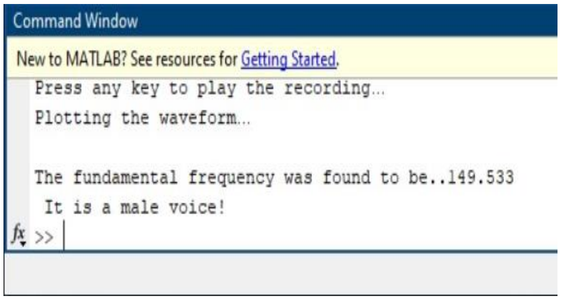


Fig 6. Autocorrelated cepstrum of the male voice.



Command Window

New to MATLAB? See resources for Getting Started.

```
Press any key to play the recording...
Plotting the waveform...

The fundamental frequency was found to be..149.533
 It is a male voice!
fx >>
```

Fig 7. Algorithm Results for a male voice.

4.2 Case 2: Female Voice:

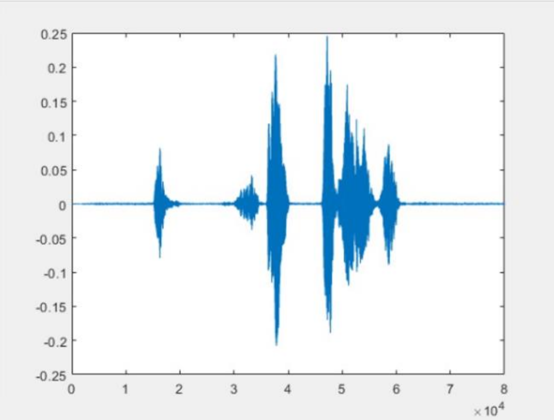


Fig 8. Raw speech signal of a female voice.

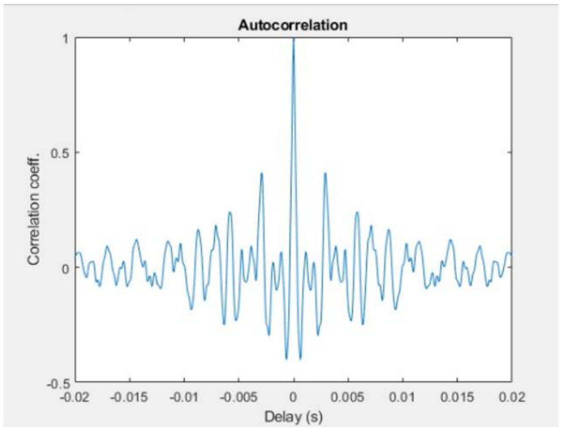
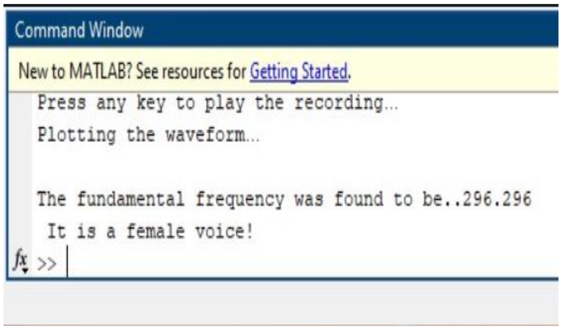


Fig 9. Autocorrelated signal of the female voice.



```
Press any key to play the recording...
Plotting the waveform...

The fundamental frequency was found to be..296.296
 It is a female voice!
>>
```

Fig 10. Algorithm Results for Female Voice.

Inference

The process of auto correlation has been applied in MATLAB and it has been used to check whether the real time input voice is a male or a female voice. The program takes in the real time speech voice and plots it initially along with the reading of the frequency of the voice. It then analyses the voice using autocorrelation process and plots the graph, it then takes the fundamental frequency of the voice it then checks with the threshold frequency which has been set at 160 hertz, hence when the fundamental frequency of the voice is above 160 hertz the voice will be considered as a female voice and vice versa. One more important observation is that when you compare the autocorrelation cepstrum of male and female voice, we cannot fail to notice that the female voice ceptrum has more peaks than the male counterpart. This observation reiterates the fact that female voices have higher pitch than male voices.

V. CONCLUSION

Hence, we have successfully implemented a signal processing technique called Correlation to determine the fundamental frequency of a raw speech signal. We receive the raw speech signal from the user in real-time and determine the fundamental frequency. As we can see from the results above that the frequency of the voice was first taken and then plotted and when the fundamental frequency of the voice was greater than 160 hertz it identifies the voice as female at the same time, we can also see that when the fundamental frequency was lesser than 160 hertz it identified it to be a male voice using autocorrelation. Hence, we can conclude that autocorrelation can be used to identify gender based on voice analysis.

REFERENCES

[1] Oppenheim, A., Willsky, A., & Hamid, W. (1996). Signals and Systems (2nd ed.). Pearson.

[2] Sundararajan, D. (2008). A Practical Approach to Signals and Systems. Wiley.

[3] Lathi, B. P., & Green, R. (2017). Linear Systems and Signals (The Oxford Series in Electrical and Computer Engineering) (3rd ed.). Oxford University Press.

[4] M. Kumari and I. Ali, "An efficient algorithm for Gender Detection using voice samples," 2015 Communication, Control and Intelligent Systems (CCIS), Mathura, India, 2015, pp. 221-226, doi: 10.1109/CCIntelS.2015.7437912.

[5] S. Furui, "Cepstral analysis technique for automatic speaker verification," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 2, pp. 254-272, April 1981, doi: 10.1109/TASSP.1981.1163530.

[6] Takagi, T., Seiyama, N. and Miyasaka, E. (2000), A method for pitch extraction of speech signals using autocorrelation functions through multiple window lengths. Electron. Comm. Jpn. Pt. III, 83: 67-79.

[7] B. Yegnanarayana and K. Sri Rama Murty, "Event-Based Instantaneous Fundamental Frequency Estimation From Speech Signals," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 4, pp. 614-624, May 2009, doi: 10.1109/TASL.2008.2012194.